Digital Preservation: An Introduction

KCAA Symposium – October 2, 2004

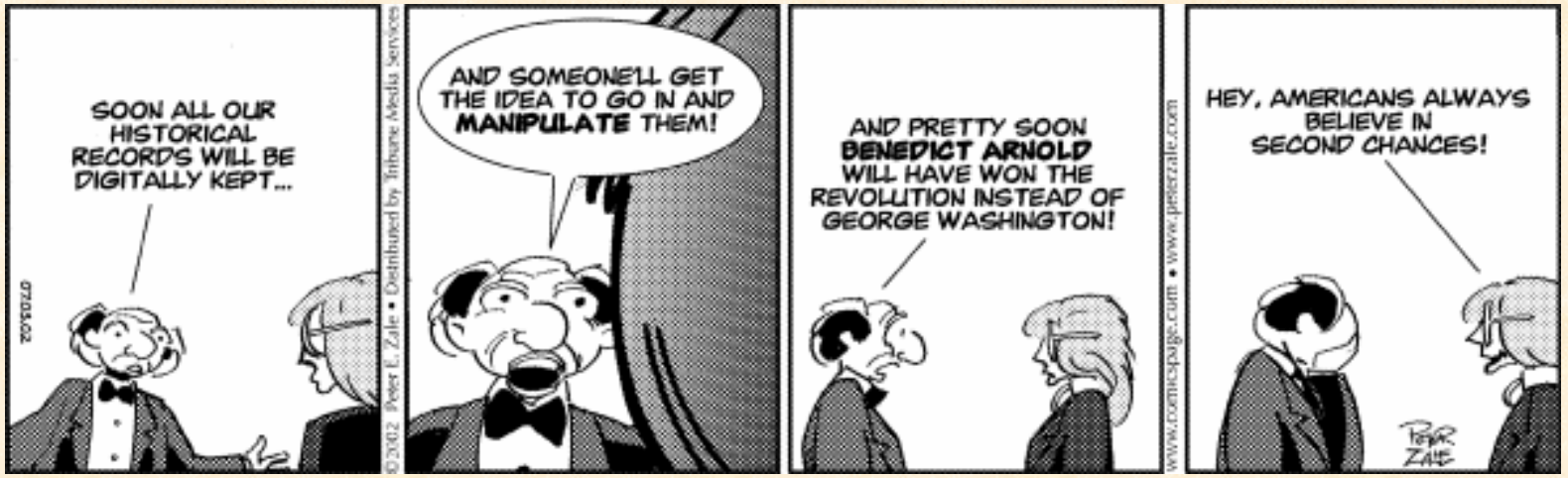
- Scott Leonard, Electronic Records Specialist
- Matt Veatch, Asst. Director, Library/Archives Division

KANSAS
STATE HISTORICAL SOCIETY

Another View

Helen, Sweetheart of the Internet

by Peter Zale (July 3, 2002)



KANSAS
STATE HISTORICAL SOCIETY

Outline

- Advantages of Digital Information
- Issues Raised by Digital Information
- Principles and Best Practices
- Overview of Current Research
- Resources
- Questions

KANSAS
STATE HISTORICAL SOCIETY

New Opportunities through the Use of Digital Technology

- Convenience of Dissemination
- Lowering of Geographical Barriers
- Quick Updating
- Rapid Feedback
- Data Manipulation and Repurposing
- Integration of Services

KANSAS
STATE HISTORICAL SOCIETY

Issues Raised by Digital Materials

- Fragile Media
- Technology Dependence
- Obsolescence
- Easy to Copy, Hard to Maintain
- Distributed Custody

KANSAS
STATE HISTORICAL SOCIETY

Principles and Best Practices

- Get Involved Early
- Partnerships
- Economic Sustainability
- Content vs. Format
- Reliability and Authenticity
- Preservation Strategies
- File Formats for Preservation
- Storage Media
- Metadata
- Digital Preservation Policy

KANSAS
STATE HISTORICAL SOCIETY

Principles and Best Practice

Get Involved Early

- Influence System Design for Recordkeeping
  - May be more practical in an institutional archives setting
    - Within a records management environment
  - Guidelines are a good tool in this area
  - Need to partner with system designers

KANSAS
STATE HISTORICAL SOCIETY

Principles and Best Practice

Partnerships

- With Records Creators
- With Technologists

KANSAS
STATE HISTORICAL SOCIETY

Principles and Best Practice

Economic Sustainability

- Tradeoffs
  - Can you maintain the material digitally forever?
  - Is there a need to keep material digitally?
  - Does the file need to remain in its native format?
- Analog Options
  - Microfilm vs. Paper

KANSAS
STATE HISTORICAL SOCIETY

Principles and Best Practice

Content vs. Format

- Make Appraisal Decisions Based on Content
  - While format will influence appraisal it should not dictate the final decision
    - E-Mail = Correspondence
- Format Will Shape Storage/Access Decisions
  - Advantage to keeping digital?
    - Ease of access, repurposing, etc.
  - Parallel in Analog Environment

KANSAS
STATE HISTORICAL SOCIETY

Principles and Best Practice

Reliability and Authenticity

- Reliability – Data Reflect Real Events and Activities (Record Is What You Say It Is)
- Authenticity – Reliability Is Maintained Over Time (Record Is Still What It Used to Be)
- Both Key to Ensuring Trustworthy Digital Information

KANSAS
STATE HISTORICAL SOCIETY

Principles and Best Practice

Elements of Reliability and Authenticity

- **Content** – that which conveys information (e.g. text, data, symbols, numerals, images, and sound)

KANSAS
STATE HISTORICAL SOCIETY

Principles and Best Practice

Elements of Reliability and Authenticity

- **Context** – background information that enhances understanding of technical and business environments to which the records relate (e.g. metadata, application software, logical business models) and the origin (e.g. address, title, link to function or activity, agency, program or section)

KANSAS
STATE HISTORICAL SOCIETY

Principles and Best Practice

Elements of Reliability and Authenticity

- **Structure** – appearance and arrangement of the content (e.g. relationships between fields, entities, language, style, fonts, page and paragraph breaks, internal links)

KANSAS
STATE HISTORICAL SOCIETY

Principles and Best Practice

Preservation Strategies

- Reformatting
  - "…an alteration in the underlying bit stream but there is no change in the representation or intellectual content of the [information]." (Charles Dollar)
  - From one storage medium to another that is different.
- Refreshing
  - a.k.a., "Copying"
  - From old storage media to new storage media with the same format specifications.
  - Essentially there is no alteration or change in the underlying bit stream.

KANSAS
STATE HISTORICAL SOCIETY

Principles and Best Practice
Preservation Strategies

- Conversion
  - Automatic exporting or importing of digital information from one software environment to another.
  - Little or no loss in structure and no loss in content or context although underlying bit stream is altered.
- Migration
  - "Moving [digital information] from one platform or program to another through the use of programming code written specifically for this purpose." (Cal Lee)

KANSAS
STATE HISTORICAL SOCIETY

Principles and Best Practices

Digital Storage Media

- No digital storage media is permanent or archival in a physical sense
- Obsolescence a bigger threat than media degradation
- Storage and handling are critical
- Periodic refreshment recommended regardless of media

KANSAS
STATE HISTORICAL SOCIETY

Principles and Best Practices
Digital Storage Media

- Optical media
  - CD-R
    - Inexpensive, high capacity, widely used
    - Susceptible to scratches, fingerprints, light, temperature, humidity, pollutants
    - Disc quality quite variable
  - DVD-R
    - Very high capacity but more expensive
    - Standards not firmly established
    - Also susceptible to environmental damage

KANSAS
STATE HISTORICAL SOCIETY

Principles and Best Practices
Digital Storage Media

- Optical media storage and handling
  - Do
    - Store in cool, dry, dark, clean-air environment
    - Store in plastic cases or protective sleeves
    - Store discs upright (book style)
    - Label the clear inner hub of the disc with a non-solvent-based marker

KANSAS
STATE HISTORICAL SOCIETY

Principles and Best Practices
Digital Storage Media

- Optical media storage and handling
  - Do not
    - Touch the surface of the disc
    - Bend the disc
    - Use adhesive labels
    - Expose disc to high temp/humidity or to rapid fluctuations in temp/humidity
    - Expose disc to UV light for long periods
    - Write on data area of disc
    - Use solvent based marker to label disc

KANSAS
STATE HISTORICAL SOCIETY

Principles and Best Practices
Digital Storage Media

- Magnetic media
  - Tape
    - Not recommended for long-term storage
    - Generally used for backups
  - Hard drives (Network servers, SANs, NAS)
    - Quicker access to digital content
    - Expensive but costs are decreasing
    - Redundancy and error checking can be built in

KANSAS
STATE HISTORICAL SOCIETY

Principles and Best Practices

File Formats for Preservation

- File = Sequence of bits
- File format = How to interpret the bits
- File format categories
  - Text (Plain text, PDF, HTML, RTF, Word)
  - Images (TIFF, JPEG, GIF)
  - Databases (Delimited text, SQL, DBF, .mdb)
  - Audio (AIFF, WAV, MP3, Real Audio)
  - Video (MPEG, AVI, Quicktime, Real Video)

KANSAS
STATE HISTORICAL SOCIETY

Principles and Best Practices

File Formats for Preservation

- Requirements for preservation
  - Full specifications publicly available
  - Recognized by a national or international standards body (ISO, ANSI, IEEE, etc.)
  - Widely used and accepted
  - Free of patent and license fees
  - Free of encryption or compression techniques
  - Platform independent

KANSAS
STATE HISTORICAL SOCIETY

Principles and Best Practices

File Formats for Preservation

- "Preferred" formats:
  - Text – Plain text, XML, PDF-A (eventually)
  - Images – TIFF
  - Databases (Delimited text, SQL DDL)
  - Audio – AIFF, WAV
  - Video – MPEG-1

KANSAS
STATE HISTORICAL SOCIETY

Principles and Best Practices

File Formats for Preservation

- In practice "acceptable" formats may have to be good enough.
  - Text – PDF, RTF
  - Images – JPEG
  - Databases – DBF

KANSAS
STATE HISTORICAL SOCIETY

Principles and Best Practices

File Formats for Preservation

- But try to avoid formats that meet few if any of preservation requirements
  - Text – Word, WordPerfect
  - Images – GIF, PhotoShop
  - Database – Access (.mdb)
  - Audio – MP3, Real Audio
  - Video – AVI, Quicktime, Real Video

KANSAS
STATE HISTORICAL SOCIETY

Principles and Best Practices

Metadata

- Metadata = Data about data
- Descriptive metadata
  - Information about the content of the digital object
    - Author, title, date created, keywords
    - Facilitates information retrieval
- Structural metadata
  - Describes the internal structure of digital objects and the relationships between their parts
    - A book is composed of chapters and chapters are composed of pages.

KANSAS
STATE HISTORICAL SOCIETY

Principles and Best Practices
Metadata

- Administrative metadata
  - Information used to manage a digital object
    - Critical for digital preservation
  - Preservation metadata a subset
    - Creation method
    - File format
    - Migration history
    - Hardware/software dependencies

KANSAS
STATE HISTORICAL SOCIETY

Principles and Best Practices

Metadata

- Metadata standards
  - Dublin Core
    - 15 elements: Title, Creator, Subject, Description, Date . . .
    - www.dublincore.org
  - METS (Metadata Encoding & Transmission Standard)
    - XML-based standard for describing complex digital objects
    - Provides a way to identify the digital pieces that together comprise a digital object, for specifying the location of these pieces, and for expressing the structural relationships between them.
    - http://www.loc.gov/standards/mets/
  - MARC
  - MODS (Metadata Object Description Schema)
    - XML-based simplified version of MARC
    - http://www.loc.gov/standards/mods/

KANSAS
STATE HISTORICAL SOCIETY

Principles and Best Practices
Digital Preservation Policy

- Putting it all together in a formal policy
- Content guidelines
  - What digital assets warrant preservation?
- Preservation strategies
  - File formats
  - Metadata
  - Storage media
  - Refreshing, conversion, migration, emulation
  - Technology forecasting

KANSAS
STATE HISTORICAL SOCIETY

Overview of Current Research

Preservation of Web Sites

- Internet Archive
  - www.archive.org/
  - "…building a digital library of Internet sites and other cultural artifacts in digital form."
- WI Historical Society
  - *Archival Issues* 27, no. 2
  - "The Governor's Blue-Ribbon Commission on State-Local Partnerships for the 21st Century" a.k.a. The Kettl Commission
  - Received a CD-ROM Copy of the Commission Web Site

KANSAS
STATE HISTORICAL SOCIETY

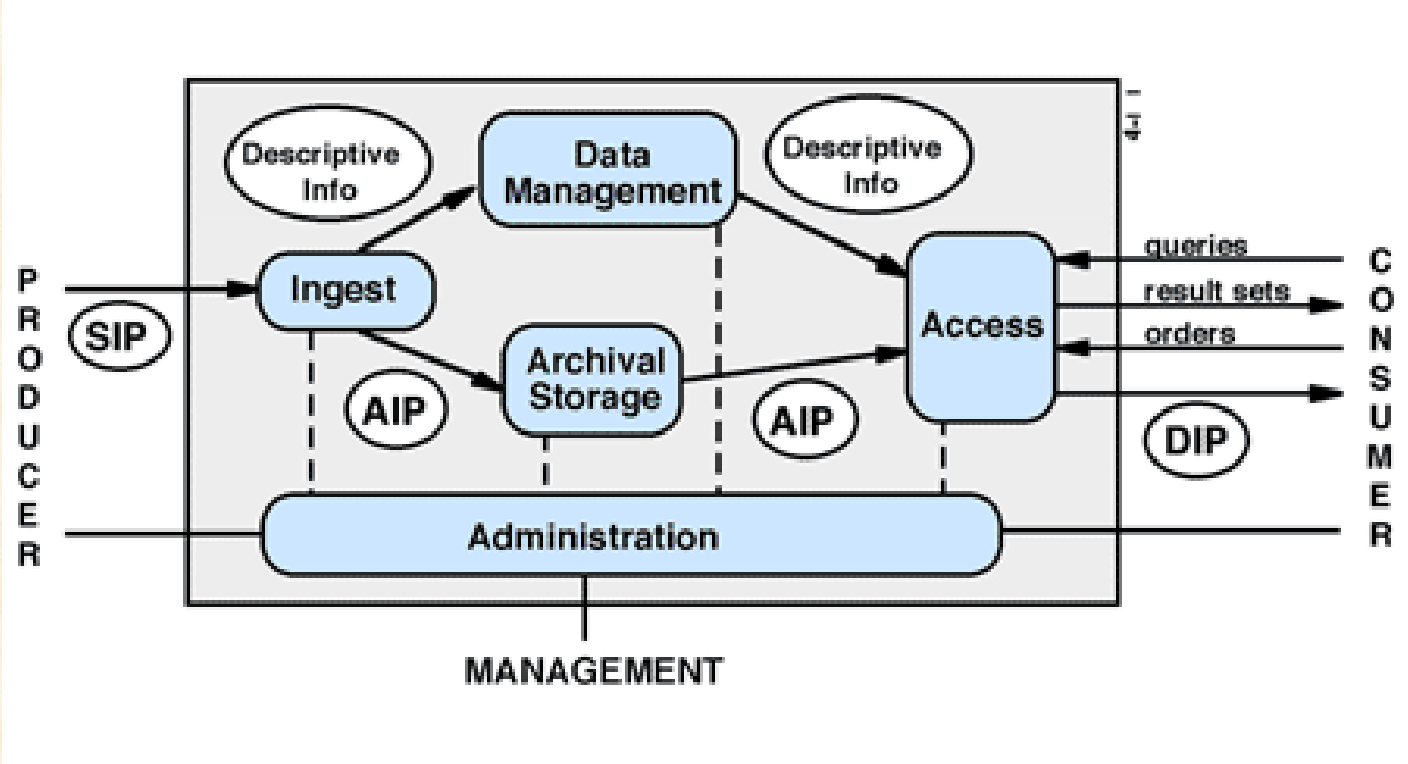
Overview of Current Research
OAIS Reference Model

- Open Archival Information System
  - ISO Standard
  - Provides, among other things,:
    - A framework for the understanding and increased awareness of archival concepts needed for long term digital information preservation and access;
    - A framework for describing and comparing architectures and operations of existing and future archives.
    - A framework for describing and comparing different long term preservation strategies and techniques;
  - www.rlg.org/en/page.php?Page_ID=3201&projGo.x=18&projGo.y=14

KANSAS
STATE HISTORICAL SOCIETY

Overview of Current Research

OAIS Reference Model



KANSAS
STATE HISTORICAL SOCIETY

Overview of Current Research

Electronic Records Archive

- National Archives and Records Administration
  - www.archives.gov/electronic_records_archives/
  - "A comprehensive, systematic, and dynamic means for preserving virtually any kind of electronic record, free from dependence on any specific hardware or software."
  - Intended to be Scaleable
  - Project Launched in 1998
  - Five Years of Preliminary Research
  - $20.1 Million Contract Awarded
    - Lockheed Martin
    - Harris Corporation

KANSAS
STATE HISTORICAL SOCIETY

Overview of Current Research

Persistent Archives Testbed

- San Diego Supercomputer Center and Others
  - www.sdsc.edu/PAT
  - "Community model for electronic records management, with archival and technological functions practically and appropriately allocated in a distributed network."
  - Automate Archival Processes Including:
    - Appraisal
    - Accessioning
    - Arrangement
    - Description
    - Preservation
    - access

KANSAS
STATE HISTORICAL SOCIETY

Overview of Current Research

Institutional Digital Repositories

- DSpace
  - www.dspace.org
  - Developed Jointly by MIT Libraries and Hewlett-Packard
  - Developed to Capture, Preserve and Provide Access to University Electronic Research and Publications
  - KSPACe

KANSAS
STATE HISTORICAL SOCIETY

Overview of Current Research

Institutional Digital Repositories

- Fedora™
  - www.fedora.info/
  - Developed Jointly by the University of Virginia and Cornell University
  - Can be used to support: institutional repositories, digital libraries, content management, digital asset management, scholarly publishing, and digital preservation
  - Tufts/Yale Project

KANSAS
STATE HISTORICAL SOCIETY

Overview of Current Research
LOCKSS

- "Lots of Copies Keeps Stuff Safe"
  - lockss.stanford.edu/
  - "Creates low-cost, persistent digital "caches" of authoritative versions of http-delivered content."
  - Similar to print model.
  - Distribute copies of digital content across multiple repositories.
  - If one copy is destroyed, other copies available to replace it.

KANSAS
STATE HISTORICAL SOCIETY

Overview of Current Research
DAITSS

- Dark Archive in the Sunshine State
  - www.fcla.edu/digitalArchive/daInfo.htm
  - Florida Center for Library Automation (FCLA) creating OAIS-based digital archives for the libraries of the public universities of Florida.
  - File format work particularly useful

KANSAS
STATE HISTORICAL SOCIETY

Overview of Current Research

InterPARES

- International Research on Permanent Authentic Records in Electronic Systems
- InterPARES 1
  - www.interpares.org
  - Authenticity of non-current electronic records designated for permanent preservation.
    - What is required to prove the authenticity of electronic records?
    - How do we select electronic records for preservation?
    - How do we preserve authentic electronic records?
    - What policies, strategies, and standards will protect the authenticity of electronic records over time?

KANSAS
STATE HISTORICAL SOCIETY

Overview of Current Research

InterPARES

- InterPARES 2
  - www.interpares.org
  - Research on preserving records produced by emerging dynamic and interactive technologies

KANSAS
STATE HISTORICAL SOCIETY

Overview of Current Research
CEDARS

- CURL (Consortium of University Libraries) Exemplars in Digital Archives
  - www.leeds.ac.uk/cedars/index.html
  - Developed digital preservation best practices
    - Preservation metadata
    - Digital collections management
    - Intellectual property rights

KANSAS
STATE HISTORICAL SOCIETY

Overview of Current Research
CAMiLEON

- Creative Archiving at Michigan & Leeds: Emulating the Old on the New
  - www.si.umich.edu/CAMILEON/
  - Examined technology emulation as a digital preservation strategy

KANSAS
STATE HISTORICAL SOCIETY

Conclusions

- Get Involved Early
- Partnerships
- Learn From Others
  - Monitor the Research
  - University, Government and Business
  - Nationally
    - NARA
    - Other States
  - Internationally
    - UK
    - Australia

KANSAS
STATE HISTORICAL SOCIETY

Resources – Web Sites

- Electronic Records Management Web Site (KSHS)
  - www.kshs.org/government/records/electronic/index.htm
- National Digital Information Infrastructure and Preservation Program (Library of Congress)
  - www.digitalpreservation.gov/index.php
- Electronic Resource Preservation and Network (ERPANET)
  - www.erpanet.org/
- Digital Preservation: Best Practice for Museums – Bibliography (Canada Heritage)
  - www.chin.gc.ca/English/Digital_Content/Digital_Preservation/bibliography.html

KANSAS
STATE HISTORICAL SOCIETY

Resources – Literature

- *Authentic Electronic Records: Strategies for Long-Term Access* (Dollar, Cohasset Associates, Inc., 1999)
- "Thirteen Ways of Looking at … Digital Preservation" (Lavoie & Dempsey, *D-Lib Magazine* 10, no. 7/8)
  - www.dlib.org/dlib/july04/lavoie/o7lavoie.html
- "Waiting for the Ghost Train: Strategies for Managing Electronic Personal Records Before It Is Too Late" (Cunningham, *Archival Issues* 24, no. 1)

KANSAS
STATE HISTORICAL SOCIETY

Questions?

- Scott Leonard
  - *sleonard@kshs.org*
  - (785) 272-8681 ext. 280
- Matt Veatch
  - *mveatch@kshs.org*
  - (785) 272-8681 ext. 271

KANSAS
STATE HISTORICAL SOCIETY